



## The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment

Hassager, Henrik Gert; Gran, Fredrik; Dau, Torsten

*Published in:*  
Journal of the Acoustical Society of America

*Link to article, DOI:*  
[10.1121/1.4950847](https://doi.org/10.1121/1.4950847)

*Publication date:*  
2016

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Hassager, H. G., Gran, F., & Dau, T. (2016). The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. *Journal of the Acoustical Society of America*, 139(5), 2992-3000. <https://doi.org/10.1121/1.4950847>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment

Henrik Gert Hassager

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark,  
DK-2800 Kongens Lyngby, Denmark

Fredrik Gran

Research Group, GN ReSound A/S, DK-2750 Ballerup, Denmark

Torsten Dau<sup>a)</sup>

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark,  
DK-2800 Kongens Lyngby, Denmark

(Received 18 June 2015; revised 3 May 2016; accepted 4 May 2016; published online 25 May 2016)

Individual binaural room impulse responses (BRIRs) were recorded at a distance of 1.5 m for azimuth angles of 0° and 50° in a reverberant room. Spectral details were reduced in either the direct or the reverberant part of the BRIRs by averaging the magnitude responses with band-pass filters. For various filter bandwidths, the modified BRIRs were convolved with broadband noise and listeners judged the perceived position of the noise when virtualized over headphones. Only reductions in spectral details of the direct part obtained with filter bandwidths broader than one equivalent rectangular bandwidth affected externalization. Reductions in spectral details of the reverberant part had only little influence on externalization. In both conditions, externalization was not as pronounced at 0° as at 50°. To characterize the auditory processes that may be involved in the perception of externalization, a quantitative model is proposed. The model includes an echo-suppression mechanism, a filterbank describing the frequency selectivity in the cochlea and a binaural stage that measures the deviations of the interaural level differences between the considered input and the unmodified input. These deviations, integrated across frequency, are then mapped to a value that corresponds to the perceived externalization. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4950847>]

[FJG]

Pages: 2992–3000

## I. INTRODUCTION

One fascinating aspect of the human hearing sense is its ability to capture the surrounding auditory space from the often complex acoustic input to the two ears. Even in reverberant environments and in the presence of multiple sound sources, the auditory system is able to extract the relevant acoustic information provided by the binaural room impulse responses (BRIRs), such that sound sources are perceived as *externalized* and as arising from the reverberant environment. The BRIRs represent the impulse responses from the sound source positions to the listener's left and right ears. The impulse responses consist of the filtering by the head, torso, and pinna, as described by the head-related transfer functions (HRTFs), and the acoustic interaction with the reflective environment in which the listener is present (Hartmann and Wittenberg, 1996; Begault *et al.*, 2001). According to Hartmann and Wittenberg (1996), externalized sound images are perceived by the listener to be compact and correctly located in space. Convincingly externalized sound images can be obtained via headphones, if the headphone reproduction includes the filtering by the BRIRs (Begault *et al.*, 2001). However, when sounds presented via headphones or other listening devices, such as hearing aids,

lack the filtering by the BRIRs, the externalization breaks down and the sound images are most likely perceived inside of the listener's head, i.e., internalized. In studies on distance perception, the ratio of the acoustic energy in the direct sound versus the reverberant part of the sound, i.e., the direct-to-reverberant ratio (DRR), has been demonstrated to relate to the perceived sound source distance (e.g., Zahorik *et al.*, 2005; Kopčo and Shinn-Cunningham, 2011). However, several studies on sound externalization suggested that the spectra of the left-ear and the right-ear stimuli affect the amount of perceived externalization (e.g., Hartmann and Wittenberg, 1996; Boyd *et al.*, 2012).

Some studies investigated the effect of modifications of the (spectral) shape of the HRTF on the spatial perception of the sound. Specifically, Kulkarni and Colburn (1998) examined how a reduction of the spectral details of individual HRTFs affects sound localization in anechoic conditions whereby no head movements were allowed but visual cues were present (Kulkarni and Colburn, 1998). In their study, a smoothing of the spectral details of the HRTF was obtained by truncating the Fourier series of the HRTF spectra. The influence of the reduced spectral details was studied for azimuth angles of a single sound source at 0°, 45°, 135°, and 180°, using sounds externalized over tube phones. By comparing virtual and real loudspeaker stimuli, Kulkarni and Colburn (1998) found that the number of coefficients in the

<sup>a)</sup>Electronic mail: tdau@elektro.dtu.dk

Fourier series could be reduced from 512 to 16 without affecting the perceived sound localization, i.e., all listeners reported fully externalized virtual sounds even in the conditions with the strongest spectral smoothing.

Breebaart and Kohlrausch (2001) investigated to what extent reductions in the spectral details of HRTFs were detectable. They smoothed the magnitude and phase spectra of nonindividual HRTFs with a gammatone filterbank (Patterson *et al.*, 1988) with different filter orders to achieve different degrees of smoothing. The fourth-order gammatone filterbank had individual filter bandwidths corresponding to the equivalent rectangular bandwidth (ERB) scale according to Glasberg and Moore (1990) which roughly reflects the frequency-selective processing in normal-hearing listeners. Listeners compared sounds that were convolved with the smoothed vs original HRTFs and were asked to rate the perceived audible difference on a three-step scale. Azimuth angles of 0°, 30°, and 120° were considered in their study and it was found that a smoothing with filter orders above one did not produce any audible effect whereas smoothing with filter orders below one was detectable.

The above studies were either carried out only in an anechoic environment (Kulkarni and Colburn, 1998), i.e., in the absence of any reverberation, or were focused on pure detectability of the spectral modifications of the HRTFs (Breebaart and Kohlrausch, 2001). However, it has been demonstrated that reverberation contributes to externalization perception (e.g., Begault *et al.*, 2001; Catic *et al.*, 2013) via the binaural cues provided by the interaural level differences (ILDs) and interaural time differences (ITDs) in a given environment. Related to the DRR, Catic *et al.* (2015) found that BRIR modifications, by either BRIR truncations or making the reverberation identical in both ears, altering the binaural cues in terms of the interaural coherence (IC) are important for sound source externalization. Thus the binaural interaction of the direct part as well as early reflections and the late reverberation of the BRIR have been found to affect externalization. However, a detailed understanding of the role of the different parts of the BRIR on externalization is still missing.

The present study investigated the influence of spectral smoothing of the BRIR on externalization in a reverberant environment. Two experiments were conducted to study the effects of early reflections and late reverberation on externalization as a function of the spectral fidelity of the processing. In the first experiment, the direct part of the BRIR was spectrally smoothed, similar as in the study of Breebaart and Kohlrausch (2001), but combined with the reverberant part of the BRIR which was left unchanged. In the second experiment, the corresponding effect of spectral smoothing of the reverberant part of the BRIR (containing the early reflections and late reverberation) was investigated whereby the direct part was left unchanged. By applying spectral smoothing only on the direct part of the BRIR, the effects of the acoustical properties of the (modified) sound source on externalization (e.g., due to spectral coloration of individual ears) were studied. When applying spectral smoothing on the reverberant part of the BRIR, the effects of the acoustical properties of delayed versions of the sound source on

externalization were studied. The experiments were carried out not allowing the listeners to move their heads, since head movements have been demonstrated to affect externalization (Brimijoin *et al.*, 2013).

Furthermore, in an attempt to characterize the auditory processes that may be involved in externalization perception, a simple quantitative model was developed. The model consisted of several stages of monaural and binaural auditory pre-processing of the input stimuli, including an echo-suppression mechanism (where the direct sound was assumed to partly suppress the lagging reverberant sound components) and a mapping between the respective internal representation of the stimuli and their corresponding externalization percept. The model was applied to the conditions of the two experiments considered in the present study.

## II. METHODS

### A. Externalization measurements

#### 1. Listeners

Seven listeners, six males and one female, with audiometric pure-tone thresholds below 20 dB hearing level between 125 and 8000 Hz, aged between 33 and 40 years, participated in the experiment. All listeners were familiar with psychoacoustic localization and externalization experiments. Prior to all experiments, training was conducted to make the listeners familiar with the different degrees of externalization they would encounter in the experiments.

#### 2. BRIR measurements

Individual BRIRs were recorded with a source-to-listener distance of 1.5 m and for azimuth angles of 0° and 50°. The recordings were carried out in a reverberant listening room designed in accordance with the IEC 268-13 standard. Figure 1 shows an illustration of the experimental setup.

The BRIRs were recorded with the Tucker-Davis Technologies RX8 system at 48 828 Hz, with two Etymotic Research ER-7C probe microphones placed in the ear canals of the listeners. The tip of each probe microphone was placed in an open dome used for behind-the-ear hearing aid. A maximum-length-sequence (MLS) of order 13, with 32 repetitions played from a Genelec 6010a loudspeaker, was used to obtain the speaker-to-ear impulse responses,  $x(t)_{brir}$ . The speaker-to-ear impulse responses are considered to be the BRIRs. The headphone-to-ear impulse responses,  $x(t)_{hpir}$ , from the HD650 Sennheiser headphones (placed on the listeners) to the probe microphones, were obtained with the same MLS immediately after the recordings of the headphone-to-ear impulse responses. To compensate for the headphones, the inverse headphone-to-ear impulse response,  $x(t)_{hpir,inv}$ , was calculated in the time domain using the Moore-Penrose pseudoinverse. The speaker-to-ear-canal impulse responses,  $x(t)_{brir}$ , were convolved with the inverse headphone-to-ear-canal impulse responses,  $x(t)_{hpir,inv}$ , to create filters for virtual external sound source generation. Due to inaccuracies in the placement of the probe microphones in the current recording setup, the BRIRs cannot be considered to be accurate at frequencies above 6000 Hz.

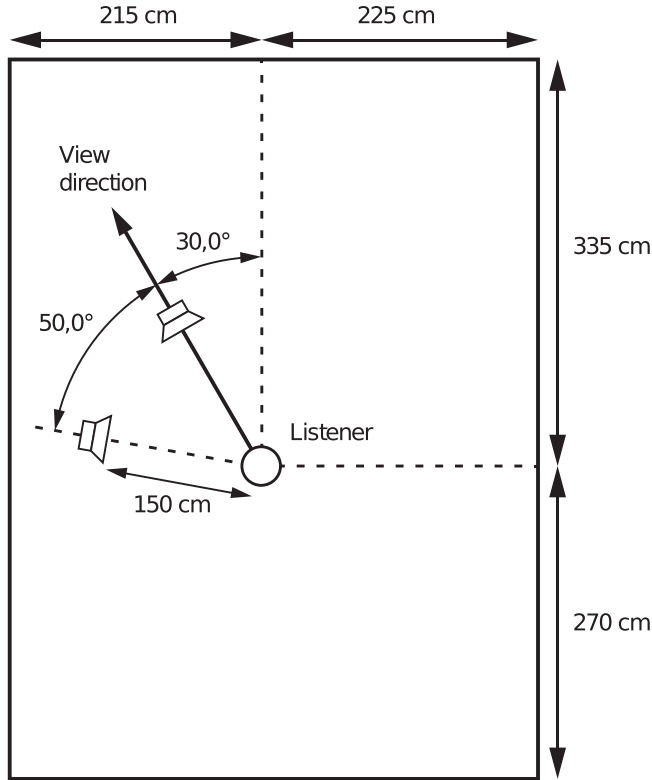


FIG. 1. The geometry of the listening room with the placement of the listener and the loudspeakers at 0° and 50°. The 0° angle is indicated by the view direction.

### 3. BRIR modifications

To study the effect of spectral detail in the BRIRs on externalization, modifications of the direct part and the reverberant part of the BRIR were undertaken. When the listeners were seated in front of the loudspeakers, the first reflection from the floor occurred after 3.8 ms. The direct part of the BRIRs was therefore defined as the first 3.8 ms, starting at the sample of most energy. The reverberant part of the BRIR after 3.8 ms comprised early reflections and late reverberation. A 5 ms half raised-cosine window was used to ensure a smooth transition of the BRIRs between the direct part and the reverberant part. Figure 2 shows the transition of the BRIR between the direct part and the reverberant part for one of the measured BRIRs.

In the time domain, the sum of the direct and the reverberant part of the BRIRs is denoted as

$$x(t)_{brir} = x(t)_{dir} + x(t)_{reverb}. \quad (1)$$

The magnitude spectrum of either the direct part or the reverberant part of the BRIRs was smoothed with a gammatone filterbank. The method described in the following is similar to the procedure used by Breebaart and Kohlrausch (2001). However, in contrast to Breebaart and Kohlrausch, the different degrees of smoothing were achieved here for different bandwidths,  $b$ , of the gammatone filters rather than different orders,  $n$ . The smoothed magnitude spectrum,  $|Y(f)|$ , can be found by calculating the smoothed frequency bins for each center frequency,  $f_c$ , given by

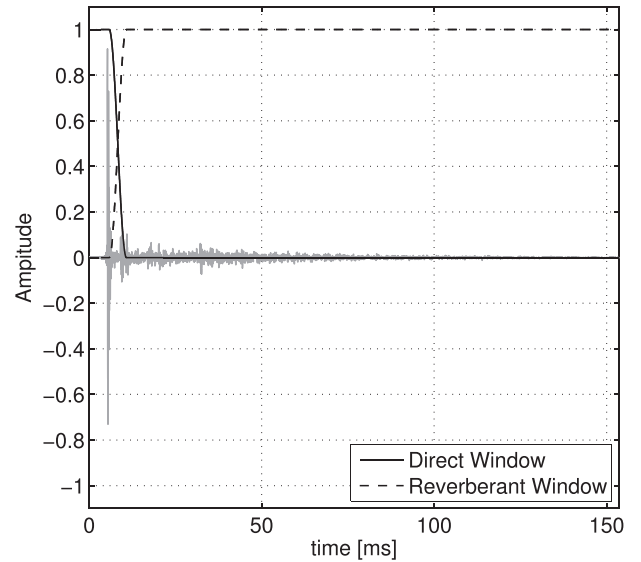


FIG. 2. The black solid and dashed functions indicate the transition windows applied on the BRIRs (in grey) to divide the BRIRs into the direct part and the reverberant part.

$$|Y(f_c)| = \frac{\sqrt{\int_0^\infty |X(f)|^2 |H(f, f_c)|^2 df}}{\sqrt{\int_0^\infty |H(f, f_c)|^2 df}} \quad (2)$$

with  $|X(f)|$  representing the original magnitude spectrum of either the direct part or reverberant part of the BRIR,  $x(t)_{dir}$  or  $x(t)_{reverb}$ , and  $|H(f, f_c)|$  denoting the magnitude spectrum of the gammatone filter at the center frequency,  $f_c$ . The approximation of the transfer function of a fourth-order gammatone filter is given as

$$H(f, f_c) = \left( \frac{1}{1 + j(f - f_c)/b(f_c)} \right)^4. \quad (3)$$

To achieve magnitude smoothing of different degrees, the bandwidth  $b(f_c)$  of the gammatone filters was represented as

$$b(f_c) = B \times 1.149 \times 24.7 \left( \frac{4.37 \times f_c}{1000} + 1 \right) \quad (4)$$

with  $B$  denoting the bandwidth factor relative to a value of one, representing the original ERB values according to Glasberg and Moore (1990).

Since it has been shown that a minimum-phase version of the HRTF is a perceptually valid description (Kulkarni et al., 1995) for the HRTF, the magnitude smoothing of the direct part of the BRIRs was applied to minimum-phase versions of the direct parts of the BRIRs. The direct part of the BRIRs was therefore decomposed into a minimum-phase filter and an all-pass filter

$$X(f)_{dir} = |X(f)|_{dir} e^{j\varphi(f)_{dir,mp}} e^{j\varphi(f)_{dir,ap}}, \quad (5)$$

where  $\varphi(f)_{dir,mp}$  and  $\varphi(f)_{dir,ap}$  indicate the phases of the minimum-phase and the all-pass filters, respectively. The all-pass filters can be considered as pure delays. The magnitude

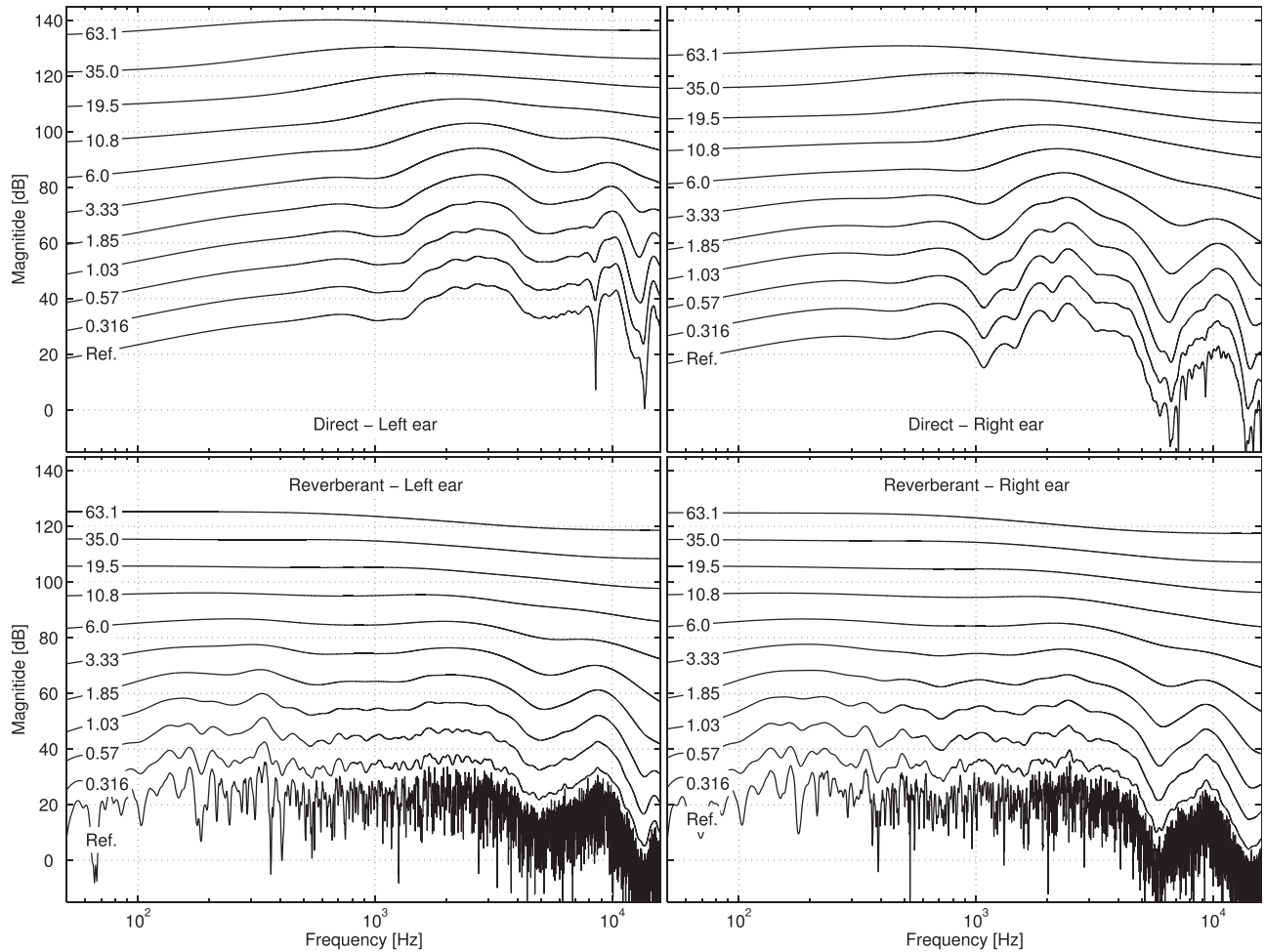


FIG. 3. Examples of original (bottom-most curves) and smoothed spectra (separated by 10 dB for each degree of smoothing for better visibility) for both the direct part (top panels) and the reverberant part (bottom panels) of the BRIR for the 50° conditions for one of the listeners. The results for the left-ear stimuli are shown in the left panels and the corresponding results for the right-ear signals are shown in the right panels.

spectra of the minimum-phase filters,  $|X(f)|_{dir}$ , were smoothed according to Eq. (2). The phase of magnitude-smoothed filters were turned into minimum-phase to ensure that the phase modifications were kept small. The magnitude smoothed minimum-phase filters were then convolved with the corresponding all-pass filters to generate the modified direct parts

$$Y(f)_{dir} = |Y(f)|_{dir} e^{j\phi(f)_{dir,mp}} e^{j\phi(f)_{dir,ap}}, \quad (6)$$

where  $\phi(f)_{dir,mp}$  denotes the phase for the magnitude smoothed minimum-phase filter. The corresponding unmodified reverberant part was then added to generate the BRIRs with the modified direct part to form BRIRs with magnitude smoothed direct parts

$$y(t)_{brir,dir} = y(t)_{dir} + x(t)_{reverb}. \quad (7)$$

The magnitude smoothing of the reverberant part of the BRIRs was applied on the magnitude response of the short-time Fourier transformation, similar to the smoothing done by Baer and Moore (1993). A short-time Fourier transformation of the reverberation part of the BRIRs was applied with an 8192-samples long Hanning window, corresponding

to 160 ms, and a step size of one sample. For each window, frequency spectra were obtained,

$$X(f)_{reverb} = |X(f)|_{reverb} e^{j\phi(f)_{reverb}}. \quad (8)$$

The magnitude spectra were smoothed, according to Eq. (2), while keeping the phase spectra unmodified:

$$Y(f)_{reverb} = |Y(f)|_{reverb} e^{j\phi(f)_{reverb}}, \quad (9)$$

where  $|X(f)|_{reverb}$  indicates the original magnitude spectrum,  $|Y(f)|_{reverb}$  indicates the smoothed magnitude spectrum, and  $\phi(f)_{reverb}$  represents the phase. The smoothed reverberant part of the BRIRs was generated via the corresponding inverse short-time Fourier transformation. The unmodified direct parts were then added to generate the BRIRs with the modified reverberant parts

$$y(t)_{brir,reverb} = x(t)_{dir} + y(t)_{reverb}. \quad (10)$$

Figure 3 shows the original spectra and the smoothed spectra, for one of the listeners, for both the direct part (top panels) and the reverberant part (bottom panels) of the BRIR for the 50° conditions. The left panels show the results for the



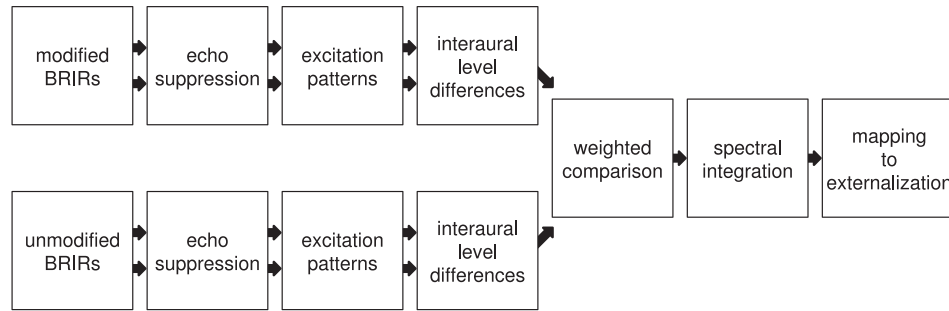


FIG. 4. Structure of the computational sound externalization model considered in the present study. The input to the model is the noise stimulus convolved with the unmodified BRIR (lower path) and the modified BRIR (upper path). This signal is processed through several stages, including an echo-suppression mechanism, a gammatone filterbank to calculate the excitation pattern at the level of the cochlea and a binaural stage that calculates the ILDs. The deviations between the modified and unmodified ILDs are calculated at each center frequency, weighted and integrated across frequency. The deviations from the reference are then mapped to the perceived externalization.

left-ear stimuli and the right panels show the corresponding results for the right-ear stimuli. The smoothed spectra were created using the same degree of smoothing as in the experiment. For better visibility, the smoothed spectra were separated by 10 dB for each degree of smoothing.

#### 4. Experimental conditions

The stimuli used for the experiments were band-pass filtered noises (50–6000 Hz) with a duration of 4 s and a sound pressure level (SPL) of 75 dB. The long duration of the stimuli was chosen to ensure that the listeners could provide a reliable response associated with their percept. BRIRs with either the modified direct part or the modified reverberant part were created with bandwidth factors,  $B$ , of 0.316, 0.570, 1.03, 1.85, 3.33, 6.0, 10.8, 19.5, 35.0, and 63.1. The noise was convolved with either the original or the modified BRIRs and then convolved with the inverse headphone-to-ear filters to create signals providing different degrees of perceived externalization of the noise. These signals were presented over the same headphones as used in the BRIR measurements.

The listeners were seated in front of the loudspeakers, which were placed at the measurement positions, as illustrated in Fig. 1. The listeners were asked to remain still and judge the perceived position of the sound source between the loudspeaker and their head. Similar to Hartmann and Wittenberg (1996), a linear scale was used to indicate the distance from the listener's head to the loudspeaker. A scale from 1 to 5 was used where 1 corresponded to a percept in the head and 5 corresponded to a percept coming from the loudspeaker. The listeners were instructed to ignore other perceptual attributes, such as frequency coloration or apparent source width, and to only focus on the degree of perceived position of the sound source. The modified signals and the original signal were each repeated 10 times with different realizations of the noise. No response feedback was provided to the listeners. The signals were presented in random order. The experiment was first carried out at the azimuth angle of  $0^\circ$  and then at the azimuth angle of  $50^\circ$ . Prior to the actual experiment, a training session was carried out, where each of the signals was played once to get the listeners familiar with the perceptual sensation. Some listeners reported changes in coloration. However, all listeners

reported the source width of the stimuli to be unchanged. Results of the experiment will be presented later in Sec. III.

#### B. Externalization model

The proposed model maps a given acoustical input signal onto a value that corresponds to the perceived externalization of the signal. The model includes several preprocessing stages, including an echo-suppression mechanism, a middle-ear filter, a gammatone filterbank describing the frequency selectivity in the cochlea, and a binaural stage that utilizes the ILD deviations of the considered stimuli from the reference ILDs of the unmodified stimuli. The deviations from the reference are then mapped to the perceived externalization. The measured BRIRs were used to calculate the reference ILDs that provide the correct externalization for the azimuth angles of  $0^\circ$  and  $50^\circ$ . It was assumed that the listener has learned the properties of the reference BRIRs, e.g., the natural ILDs given by the listener's HRTF and the statistical properties of the reverberation.

Figure 4 illustrates the individual processing steps of the model. First, a simple temporal weighting of the BRIRs was applied to both the left-ear and the right-ear impulse responses to simulate echo suppression, inspired by the processing proposed in Catic *et al.* (2015). According to Braasch *et al.* (2003), the interval of the BRIR between 4.5 and 80 ms represents the summing location window for broadband sounds. The weighting function was thus assumed to take a value of one for the portion of the BRIR up to the first reflection (3.8 ms), followed by a transition to a value of 0.01 reflecting suppression, followed by a transition from 0.01 back to one using a half raised-cosine window from 10 to 160 ms. Figure 5 shows the echo suppression window, represented on a dB scale, together with one of the measured BRIRs.

The middle-ear filtering was simulated using a 512 tap finite impulse response filter as described by Goode *et al.* (1994) and Lopez-Poveda and Meddis (2001). The frequency selectivity in both ears was calculated via the excitation pattern as a function of the filter center frequency on an ERB scale, following Glasberg and Moore (1990). Next, the ILDs (in dB) at the output of the excitation patterns were calculated for center frequencies with audible content (50–6000 Hz) both

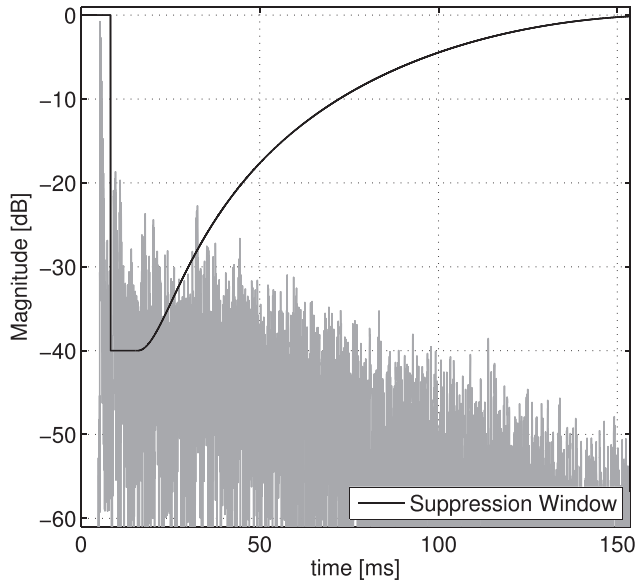


FIG. 5. The echo suppression window used in the model is indicated by the black solid function (represented on a dB scale). The grey pattern represents one of the measured BRIRs.

for the unmodified BRIRs (upper path in Fig. 4) and the modified BRIRs (lower path in Fig. 4). The deviation between the signal ILDs and the reference ILDs was calculated at each center frequency,

$$\Delta ILD = |ILD_{signal} - ILD_{reference}|. \quad (11)$$

ILD discrimination thresholds have been shown to be roughly constant at about 0.5–2 dB across frequency for a reference ILD of 0 dB (e.g., Yost and Dye, 1988). In the proposed model, deviations below 1.5 dB were therefore considered to be below threshold and set to zero,

$$\Delta ILD_{trun} = \begin{cases} \Delta ILD, & \Delta ILD \geq JND \\ 0, & \Delta ILD < JND. \end{cases} \quad (12)$$

The ratio of the ILD deviations (from the reference ILDs) to the reference ILDs was calculated at each center frequency. This implies that deviations from small reference ILDs are weighted more strongly than deviations from large references

ILDs. This is consistent with results from measurements of the minimum audible angle (e.g., Mills, 1958). An integration across the center frequencies containing audible content was undertaken to compute the overall normalized ILD deviation  $\Delta_{ILD}$ ,

$$\Delta_{ILD} = \frac{1}{N_{CF}} \sum_{i_{CF}=1}^{N_{CF}} \frac{\Delta ILD_{trun}}{|ILD_{reference}|}, \quad (13)$$

whereby it was assumed that ILDs contribute equally across frequency, as proposed by Hartmann and Wittenberg (1996). Finally, the overall normalized ILD deviation was mapped to a value corresponding to the perceived externalization of the incoming sound. The mapping is represented by a decaying exponential function

$$P_{ext} = c_1 e^{-z_1 \Delta_{ILD}} + c_2, \quad (14)$$

where  $c_1 = 3.78$ ,  $c_2 = 1$  and  $z_1 = 0.99$  represent the mapping parameters.  $c_1$  and  $z_1$  were derived from a least-squares fit and  $c_2$  was defined such that stimuli providing large ILD deviations correspond to an internalized percept ( $P_{ext} \rightarrow 1$  for  $\Delta_{ILD} \rightarrow \infty$ ). The parameters were fit only to the data obtained in the  $0^\circ$  azimuth condition where the spectral smoothing was applied to the direct sound (corresponding to a quarter of the overall data set) and were kept constant throughout the remaining conditions considered in the present study. The fitting could alternatively have been based on another subset of the experimental data without major affect on the derived parameter values.

### III. RESULTS

Figure 6 shows the mean externalization ratings (open symbols), averaged across listeners, as a function of the smoothing bandwidth factor. Simulated externalization values obtained with the proposed model are indicated by the filled symbols. The left panel shows the results obtained with the stimuli presented from the  $0^\circ$  direction whereas the corresponding results for the  $50^\circ$  azimuth are shown in the right panel.

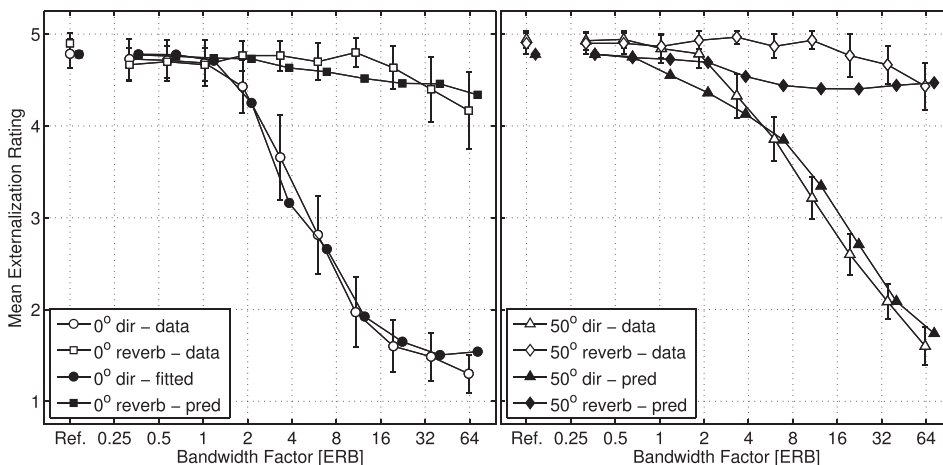


FIG. 6. The mean of the seven listeners perceived sound source location (open symbols) as a function of the bandwidth factor and the corresponding model predictions (filled symbols). The model predictions have been shifted slightly to the right for a better visual interpretation. The error bars are one standard error of the mean.

Regarding the  $0^\circ$  direction (left panel), the open circles represent the data for the condition where the direct part of the BRIRs was modified while the reverberant part was kept untouched, indicated as “ $0^\circ$  dir.” It can be seen that spectral smoothing of the direct part obtained with bandwidth factors below about 1 ERB did not, or only marginally, affect the perceptual externalization of the virtual sound source. In contrast, spectral smoothing with a bandwidth factor above 1 ERB led to decreasing externalization ratings with increasing smoothing factor such that, for the largest bandwidth factors, the sound was perceived to be close to the head or almost fully internalized. In contrast, when the reverberant part of the BRIR was spectrally smoothed whereas the direct part was kept untouched, as indicated by the open squares and referred to as  $0^\circ$  reverb in the figure, essentially no effect of the amount of the spectral smoothing on externalization was observed, i.e., the sound was always perceived as being externalized closely to the loudspeaker.

For the  $50^\circ$  direction (right panel), the overall pattern of the results was similar as in the  $0^\circ$  conditions: The externalization ratings decreased monotonically with increasing bandwidth factor above 1 ERB when the direct part was modified and the reverberant part was left unchanged, as indicated by the open triangles and referred to as the “ $50^\circ$  dir” condition. In the “ $50^\circ$  reverb” condition (open diamonds), where the reverberant part of the BRIR was modified but not the direct part, the externalization was hardly affected by the amount of spectral smoothing. However, there are also differences in the results obtained for the two source directions. In particular, for bandwidth factors above 2, the ratings obtained in the condition “ $50^\circ$  dir” were above those obtained in the condition  $0^\circ$  dir, i.e., the decay of externalization with increasing bandwidth factor was more gradual for the  $50^\circ$  than for the  $0^\circ$  condition and did not reach the same low values of externalization even for the largest bandwidth factors. In all considered experimental conditions, i.e., the conditions  $0^\circ$  dir,  $0^\circ$  reverb, “ $50^\circ$  dir,” and “ $50^\circ$  reverb,” the listeners reported consistently to perceive compact (non-diffuse) sound sources.

The simulations (filled symbols in Fig. 6) agree reasonably well with the measured data. The model describes the main effects of spectral smoothing of the BRIR on externalization both regarding the effects of modifications of the direct vs the reverberant part as well as regarding the sound source location. The simulations obtained in the (non-fitted) conditions show some deviations from the data for the intermediate bandwidth factors. However, the deviations never exceed half an externalization category.

Figure 7 shows the relationship between the measured externalization ratings (replotted from Fig. 6) and the model output after the preprocessing stages [Eq. (13)] and before the final mapping to perceived externalization [Eq. (14)]. Thus, the measured externalization ratings obtained in all conditions are now represented as a function of the overall normalized ILD deviation from the reference ILD (as defined in Sec. II B) for the respective stimulus conditions  $0^\circ$  dir (circles),  $0^\circ$  reverb (squares),  $50^\circ$  dir (triangles), and  $50^\circ$  reverb (diamonds). It can be seen that stimuli from

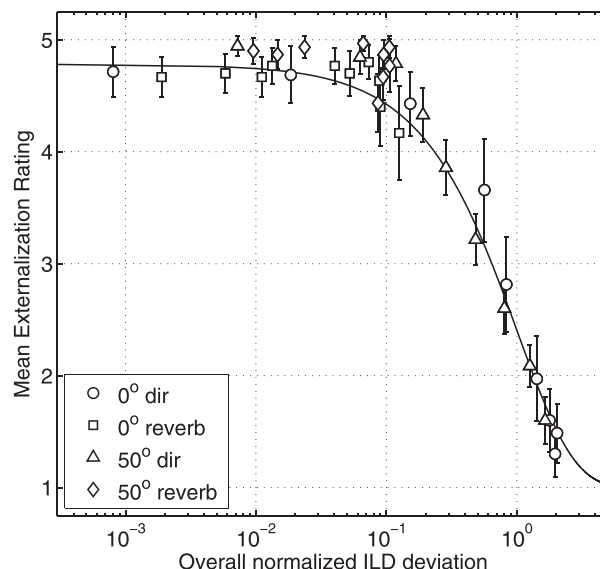


FIG. 7. The perceived sound source location as function of the overall normalized ILD deviation together with the exponential fit representing the mapping function. The overall normalized ILD deviation ( $x$  axis) is shown on a logarithm scale to display the data point linearly. The error bars are one standard error of the mean.

different experimental conditions that produced a similar externalization percept in the measurements exhibit similar values of the integrated ILD deviation. The solid function in Fig. 7 shows the least squares fit of the externalization ratings to the corresponding overall normalized ILD deviation and represents the mapping function [Eq. (14)] used in the final model step.

#### IV. DISCUSSION

The results from the present study showed that, when visual cues are provided and no head movements are allowed, the spectral details of the direct part of the BRIR contribute to sound externalization in a reverberant environment. These observations seem not consistent with the informal observations in the anechoic condition by Kulkarni and Colburn (1998), where all listeners reported fully externalized virtual sounds. However, the finding that spectral alterations affect the perceived externalization is in agreement with the results obtained in Hartmann and Wittenberg (1996) and Boyd *et al.* (2012). In the presence of early reflections and late reverberation, spectral modifications in the direct part of the BRIR (corresponding to the HRTF in an anechoic environment) are important for sound externalization in reverberant listening environments.

In studies on distance perception in reverberant environments, it has been demonstrated that the DRR of the stimuli represents an informative indicator, such that the distance of a sound source is perceived to be further away with decreasing amount of DRR (e.g., Zahorik *et al.*, 2005; Kopčo and Shinn-Cunningham, 2011). In order to test whether this metric also successfully accounts for the externalization data obtained in the present study, DRRs were calculated by convolving the direct and reverberant parts of



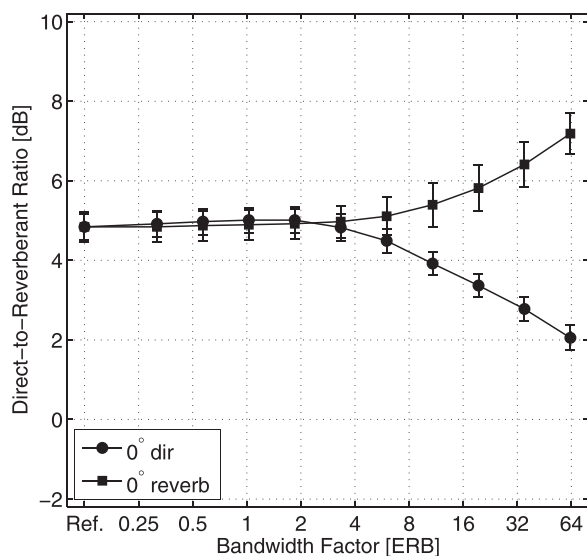


FIG. 8. The mean DRR of the left ear of the listeners for frontal conditions plotted as a function of the bandwidth factor. The error bars are one standard error of the mean.

the individual BRIRs with the noise stimuli and by computing the total power contained in the two parts, similar to [Kopčo and Shinn-Cunningham \(2011\)](#). Figure 8 shows the mean DRR for the modified BRIRs.<sup>1</sup> The mean DRRs are shown for the listeners' left-ear stimuli for the two frontal conditions, 0° dir and 0° reverb. Similar characteristics would be observed in the 50° conditions (not shown). In the condition 0° reverb (squares), the DRR increases with increasing amount of smoothing, suggesting a decrease of the perceived distance. This would be qualitatively consistent with the (slightly) decreasing amount of externalization in the data (open squares in Fig. 6). In the condition 0° dir (circles), the DRR decreases with increasing amount of smoothing, suggesting an *increase* of the perceived distance according to [Kopčo and Shinn-Cunningham \(2011\)](#). In this condition, however, the data from the present study (open circle in Fig. 6) indeed demonstrated a less externalized sound image with increasing amount of smoothing. Thus, the DRR metric does not seem to account for the conditions where the smoothing was applied to the direct part of the BRIR. Spectral modifications of the HRTF (i.e., the direct part of the BRIR) introduce deviations of the natural ILDs.

The model proposed in the present study showed that the deviations of the natural ILDs could account for the internalization of the sound stimuli, suggesting that binaural cues are crucial for correct sound externalization. However, binaural cues may not be important for robust distance perception, which may be mainly driven by monaural cues, as represented in the DRR.

To illustrate the effect of some of the processing stages of the model on simulated externalization, Fig. 9 shows the results obtained with two modifications in the model. The left panel shows simulations where the *absolute* deviations from the reference ILDs were considered, instead of the relative deviations as defined in Eq. (13). For the stimulus conditions 0° dir (circles) and 0° reverb (squares), the model no longer produces simulated externalization values along a single function (as in Fig. 7 obtained with the original model). Particularly for the spectral smoothing conditions that led to medium and small externalization ratings in the data, the simulated values for the 0° dir and the 50° dir conditions deviated strongly from each other, in contrast to the findings in the data and the original model. Thus, with this modified model version, the reduced externalization for the conditions at 0°, compared to the conditions at 50°, cannot be accounted for. The right panel of Fig. 9 illustrates simulations when no echo suppression was considered in the model. Also for this modification, the model fails to account for the data. The simulations for the different conditions no longer follow a single function. Without echo suppression, the spectral details in the early reflections and late reverberation are given too much weight and, thus, the model underestimates the amount of externalization in the conditions 0° reverb and 50° reverb, i.e., predicts more internalized values than observed in the data (and in the original model).

The presented model can only be considered as a first step towards a computational model of perceptual externalization. The assumptions in the model regarding echo suppression are qualitative and conceptual and also the processes describing frequency selectivity, excitation pattern calculation, ILD comparison, and integration across frequency have been pragmatic choices. While the individual steps have been inspired by existing auditory modeling work and psychoacoustic data, alternative implementations might be more powerful. The concept of matching the BRIR

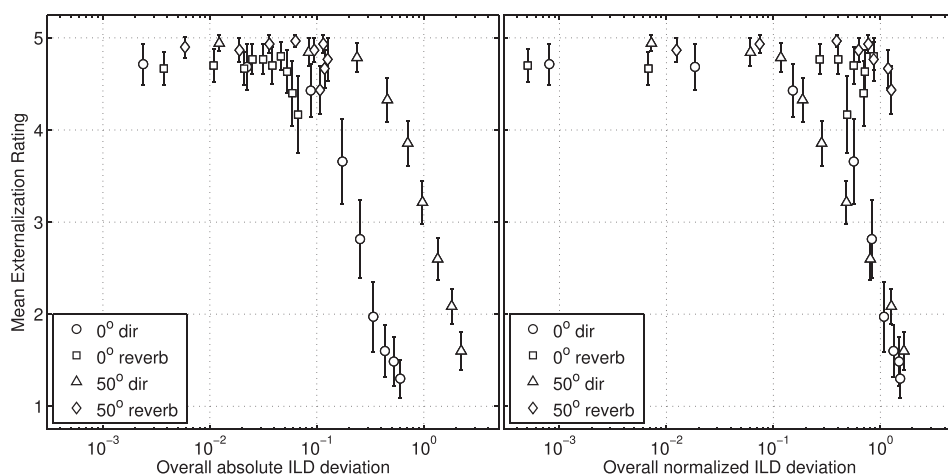


FIG. 9. The perceived sound source location as function of the overall normalized ILD deviation for the two model modifications. In the left panel the absolute deviations from the reference ILDs were considered and in the right panel shows the situation when echo suppression was omitted. The overall normalized ILD deviation (x axis) is shown on a logarithm scale to display the data point linearly. The error bars are one standard error of the mean.

to existing templates was used to estimate the amount of externalization (vs internalization). However, the template-matching method only considered ILDs and, thus, did neither take effects of ITDs nor stimulus coloration into account, which have been argued to also contribute to externalization perception (e.g., Hartmann and Wittenberg, 1996). Furthermore, dynamic changes of the convolution of BRIRs and stimuli, e.g., resulting from hearing-aid compression, are not accounted for in the model. Nevertheless, despite its simplicity, the proposed model might provide a valuable basis for further investigations of the auditory processes underlying externalization perception. Such investigations could include experimental conditions with stimuli different from those considered here, presented from different source positions and in various acoustic environments.

Changes of the BRIRs, e.g., resulting from static hearing-aid processing, will affect the pattern of ILDs (and ITDs) and might thus have an effect on the listeners' ability to externalize sound sources. Spectral modifications applied on the direct part of the BRIR will mainly affect static ILDs. For example, the position of the microphones in a hearing aid has an effect on the spectral details of the HRTF and thereby influences the static ILDs (e.g., in the case of behind-the-ear hearing aids). Thus, if the modified static ILDs due to a change of the microphone positions are not compensated for, this should affect the perceived externalization of the sounds. Less critical should be situations where fluctuating ILDs are affected, e.g., due to spectral modifications applied on the reverberant part of the BRIR (but not the direct part).

## V. SUMMARY AND CONCLUSION

In the present study, the effect of manipulating the spectral details in the binaural transfer function on perceived externalization in a reverberant room was investigated. The spectral details were reduced in either the direct or reverberant part of the BRIRs by smoothing the magnitude responses with band-pass filters. For various filter bandwidths, the modified BRIRs were convolved with broadband noise and listeners were asked, while keeping their head still, to judge the perceived position of the noise when virtualized over headphones. The data showed that reductions of the spectral details in the direct part of the BRIR in most experimental conditions had an effect on externalization. This is different from the findings obtained with corresponding spectral manipulations in an anechoic environment where no effect of spectral smearing on externalization was found. Reductions in the spectral details of the reverberant part of the BRIR did hardly affect externalization. A simple computational model was presented in an attempt to account for the data from the present study, obtained for the two source positions (0° and 50° azimuth) and for all tested modifications of the BRIR. The simulations suggested that perceived externalization can be estimated based on the deviations of the ILDs for the given modified signal from those for the (unmodified) reference signal, after some stages of auditory preprocessing (including an echo-suppression mechanism). The results from the present study might be valuable for the further investigation of the auditory processes involved in externalization perception in complex acoustic environments.

## ACKNOWLEDGMENT

This project has been carried out in connection to the Centre for Applied Hearing Research (CAHR) supported by Widex, Oticon, GN Resound and the Technical University of Denmark. Thanks to Jesper Udesen (GN Resound A/S) who performed the BRIR recordings. We also thank the Associate Editor, Frederick Gallun, and the two reviewers for their helpful and constructive feedback on an earlier version of this paper.

<sup>1</sup>According to Eq. (2), the smoothing process ensures a constant power across frequency. Since the frequency range of the stimulus (50 to 6000 Hz) and that of the smoothing process (0 to 24 414 Hz) are different, the DRRs changed as a function of the bandwidth factor used in the smoothing process.

- Baer, T., and Moore, B. C. J. (1993). "Effects of spectral smearing on the intelligibility of sentences in noise," *J. Acoust. Soc. Am.* **94**, 1229–1241.
- Begault, D. R., Wenzel, E. M., and Anderson, M. R. (2001). "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.* **49**, 904–916.
- Boyd, A. W., Whitmer, W. M., Soraghan, J. J., and Akeroyd, M. A. (2012). "Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers," *J. Acoust. Soc. Am.* **131**, EL268–EL274.
- Braasch, J., Blauert, J., and Djelani, T. (2003). "The precedence effect for noise bursts of different bandwidths. I. Psychoacoustical data," *Acoust. Sci. Technol.* **24**, 233–241.
- Breebaart, J., and Kohlrausch, A. (2001). "The perceptual (ir)relevance of HRTF magnitude and phase spectra," in *110th AES Convention*, Preprint No. 5406, Amsterdam, The Netherlands.
- Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A. (2013). "The contribution of head movement to the externalization and internalization of sounds," *PLoS ONE* **8**(12), e83068.
- Catic, J., Santurette, S., Buchholz, J. M., Gran, F., and Dau, T. (2013). "The effect of interaural-level-difference fluctuations on the externalization of sound," *J. Acoust. Soc. Am.* **134**, 1232–1241.
- Catic, J., Santurette, S., and Dau, T. (2015). "The role of reverberation-related binaural cues in the externalization of speech," *J. Acoust. Soc. Am.* **138**, 1154–1167.
- Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Goode, R. L., Killion, M., Nakamura, K., and Nishihara, S. (1994). "New knowledge about the function of the human middle ear: Development of an improved analog model," *Am. J. Otol.* **15**, 145–154.
- Hartmann, W. M., and Wittenberg, A. (1996). "On the externalization of sound images," *J. Acoust. Soc. Am.* **99**, 3678–3688.
- Kopčo, N., and Shinn-Cunningham, B. G. (2011). "Effect of stimulus spectrum on distance perception for nearby sources," *J. Acoust. Soc. Am.* **130**, 1530–1541.
- Kulkarni, A., and Colburn, H. S. (1998). "Role of spectral detail in sound-source localization," *Nature* **396**, 747–749.
- Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1995). "On the minimum-phase approximation of head-related transfer functions," in *Proceedings of the 1995 Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE, pp. 84–87.
- Lopez-Poveda, E. A., and Meddis, R. (2001). "A human nonlinear cochlear filterbank," *J. Acoust. Soc. Am.* **110**, 3107–3118.
- Mills, A. (1958). "On the minimum audible angle," *J. Acoust. Soc. Am.* **30**, 237–246.
- Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988). "SVOS final report (Part A): The auditory filterbank," APU report 2341.
- Yost, W. A., and Dye, R. H., Jr. (1988). "Discrimination of interaural differences of level as a function of frequency," *J. Acoust. Soc. Am.* **83**, 1846–1851.
- Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). "Auditory distance perception in humans: A summary of past and present research," *Acta Acust. Acust.* **91**, 409–420.